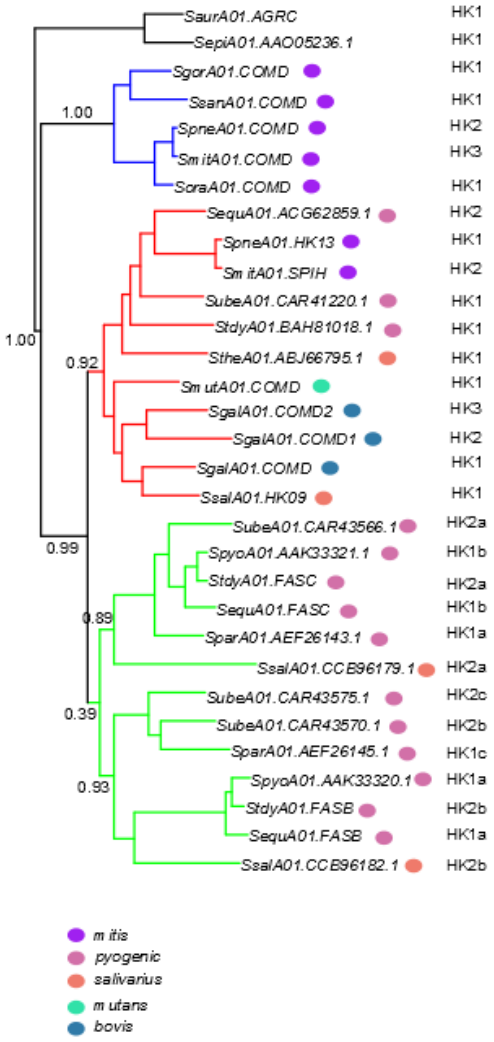
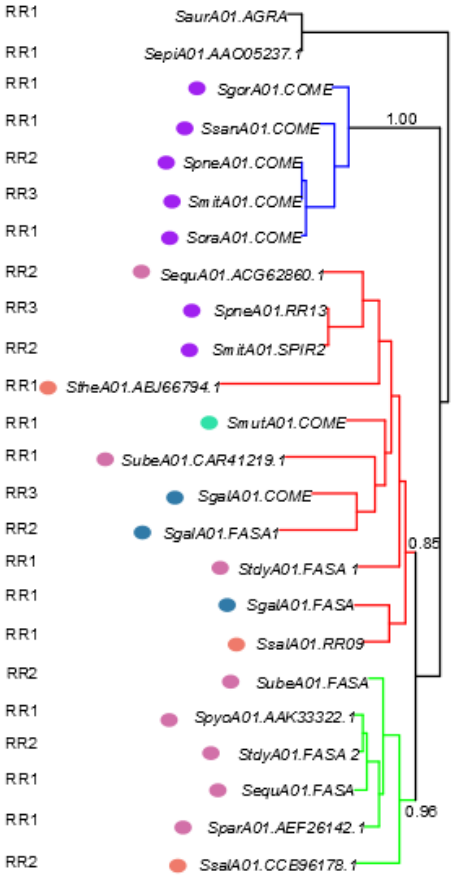
ComD


SaurA01.AGRC HK1
SepiA01.AAO05236.1 HK1
SgorA01.COMD HK1
SsanA01.COMD HK1
SpneA01.COMD HK2
SmitA01.COMD HK3
SoraA01.COMD HK1
SequA01.ACG62859.1 HK2
SpneA01.HK13 HK1
SmitA01.SPIH HK2
SubeA01.CAR41220.1 HK1
StdyA01.BAH81018.1 HK1
StheA01.ABJ66795.1 HK1
SmutA01.COMD HK1
SgalA01.COMD2 HK3
SgalA01.COMD1 HK2
SgalA01.COMD HK1
SsalA01.HK09 HK1
SubeA01.CAR43566.1 HK2a
SpyoA01.AAK33321.1 HK1b
StdyA01.FASC HK2a
SequA01.FASC HK1b
SparA01.AEF26143.1 HK1a
SsalA01.CCB96179.1 HK2a
SubeA01.CAR43575.1 HK2c
SubeA01.CAR43570.1 HK2b
SparA01.AEF26145.1 HK1c
SpyoA01.AAK33320.1 HK1a
StdyA01.FASB HK2b
SequA01.FASB HK1a
SsalA01.CCB96182.1 HK2b
1.00
1.00
0.92
0.99
0.89
0.39
0.93
mitis
pyogenic
salivarius
mutans
bovis

ComE


SaurA01.AGRA RR1
SepiA01.AAO05237.1 RR1
SgorA01.COME RR1
SsanA01.COME RR1
SpneA01.COME RR2
SmitA01.COME RR3
SoraA01.COME RR1
SequA01.ACG62860.1 RR2
SpneA01.RR13 RR3
SmitA01.SPIR2 RR2
StheA01.ABJ66794.1 RR1
SmutA01.COME RR1
SubeA01.CAR41219.1 RR1
SgalA01.COME RR3
SgalA01.FASA1 RR2
StdyA01.FASA 1 RR1
SgalA01.FASA RR1
SsalA01.RR09 RR1
SubeA01.FASA RR2
SpyoA01.AAK33322.1 RR1
StdyA01.FASA 2 RR2
SequA01.FASA RR1
SparA01.AEF26142.1 RR1
SsalA01.CCB96178.1 RR2
1.00
0.85
0.96

Les séquences utilisées pour calculer les arbres, sont similaires, homologues, orthologues et/ou paralogues ?

Les séquences ont été obtenues par une recherche par similarité dans les génomes complets et non assemblés de Streptocoques à l'aide des logiciels blastp et tblastn et en utilisant les séquences de ComD et ComE de *S. pneumoniae* comme séquences sondes (queries). Elles sont donc similaires à ComD et ComE de *S. pneumoniae* et les alignements multiples que nous avons réalisés ont montré que, par familles (ComD et ComE),les séquences étaient similaires. Nous pouvons quantifier cette similarité, par exemple, par le pourcentage d'identité entre les paires de séquences.

Nous avons fait l'hypothèse qu'elles étaient issues de gènes homologues, elles descendent d'une séquence ancêtre commune (ce sont les gènes qui sont transmis et non pas les protéines). Rappelons que des gènes peuvent être homologues sans que leurs séquences présentent une similarité significative (e-value), en raison d'un taux de mutation élevé. L'inverse est plus rare et généralement la similarité se limite à une petite portion des séquences. Cette similarité sans lien d'homologie peut être due à une convergence.

Il y a des pairs de séquences orthologues qui sont liées par un évènement de spéciation et des pairs de séquences paralogues qui sont liées par une duplication, l'identification de ces pairs nécessite une analyse détaillé des arbres (cf ci-dessous) et, en particulier, d'une localisation des évènements de spéciation et de duplication au niveau des nœuds de l'arbre.

Quel est le rôle des séquences Staphylococcus Saur et Sepi dans cet arbre?

Les séquences de Staphylococcus (Saur et Sepi) servent de groupe externe. Elles permettent d'enraciner notre arbre et donc de connaître le nœud correspondant au nœud ancêtre hypothétique de l'ensemble de nos séquences de streptocoques. Cet enracinement va nous permettre de d'ordonner l'apparition des évènements de spéciation et de duplication de gènes.

Ou placez-vous le nœud ancestral ?

Combien de groupe de séquences pouvons-nous observer sur ces deux arbres ?

Il est difficile de répondre à cette question en l'absence d'informations supplémentaires. La démarche que nous avons suivie lors de ce TP est d'identifier les groupes de feuilles qui étaient stables lors des différentes expériences réalisées avec les séquences de la famille ComE. Les groupes bleu et vert sont apparus très stables alors que le groupe rouge est identifié uniquement avec la méthode PhyML. Ces trois groupes ont des supports de bootstrap paramétrique supérieur à 0.85 avec PhyML. Les groupes bleu et rouge se retrouvent dans l'arbre ComD (bootstrap> 0.92) et le groupe vert se décompose en deux sous-groupes avec également un très bon support (>0.89) avec cependant un très faible support pour le nœud ancêtre de ces groupes (0.39).

Dans le cas de la duplication des HK du groupe vert, des signaux différents pourraient être "sentis" par chacun des senseurs et activer le même régulateur et donc intégrer deux informations différentes pour activer les mêmes gènes.

Cette distinction en trois groupes reflète des différences fonctionnelles. Les groupes bleus correspondent aux systèmes à deux composants (TCS) impliqués dans la régulation des étapes précoces de la compétence, les groupesverts renferment un TCS atypique constitué de deux histidines kinases (FasB, FasC) et d'un régulateur de réponse (FasA). L'analyse des gènes régulés par le système fas chez *S.pyogenes* suggère qu'il pourrait contrôler la bascule entre un phénotype qui favorise l'adhérence et une infection persistante et un phénotype qui conduit à la destruction des tissus et invasion rapide.Les groupes rouges correspondent à des TCS impliqués dans la production de

bactériocines de classe II. Ils appartiennent à un régulon produisant le phérormone (codé par *blpC*) et son système de sécrétion (BlpAB).

Nous avons distingué trois groupes de séquences (branches bleus, rouges et vertes).

Quelle est la distribution des espèces dans les trois groupes ?

Le groupe bleu ne renferme que des séquences du groupe mitis, le groupe vert des séquences du pyogenic et salivarius et le groupe rouge au moins un représentant de chaque groupe.

La grande distribution taxonomique observée avec le groupe rouge, avec 9 espèces représentées appartenant à 5 groupes taxonomiques (*salivarius*, *mitis*, *pyogenic*, *mutans*, *bovis*), suggère que les gènes codant pour ce système étaient présents dans l'ancêtre commun aux streptocoques et qu'un petit nombre d'espèces les auraient perdus.

Quel est le premier groupe qui émerge ?

Dans les deux arbres, ComD et ComE, à partir du nœud ancestral, le premier nœud rencontré sépare le groupe bleu des groupes rouges et verts. Ce groupe bleu renferme les séquences de ComD et ComE de de *S. pneumoniae*.

Est-ce que sa composition en espèces était attendue ?

Non, la diversité d'espèce d'un sous arbre dépend de la profondeur de l'arbre à laquelle il apparait (le dernier ancêtre commun des feuilles de ce sous arbre). Comme nous l'avons vu, le groupe rouge présente une grande diversité d'espèce suggérant qu'il est apparu très tôt au cours de l'évolution des Streptococcus. Or la topologie des arbres suggère que le groupe bleu est le premier groupe qui émerge. On s'attendait donc à ce qu'il renferme une diversité d'espèce au moins aussi grande que celle du groupe rouge qui serait apparu après lui.

Pour tenter de réconcilier ces observations (émergence précoce et diversité limité), nous pouvons faire au moins deux hypothèses. Soit, il y a effectivement une émergence précoce de ce groupe mais elle est suivit par de nombreuses pertes (délétions) dans les branches menant aux espèces n'appartenant pas au groupe mitis. Soit ce groupe a été acquis par transfert horizontal par le dernier ancêtre commun des espèces du groupe mitis. L'espèce à l'origine de ce transfert n'est pas connue mais elle était contemporaine du dernier ancêtre commun des espèces du groupe mitis.

Les différents groupes renferment des séquences issues de gènes orthologues et/ou paralogues ?

Sur l'arbre des ComE, les groupes bleu et vert ne renferment qu'une copie par génome, toutes les paires de séquences sont donc orthologues, ils constituent deux groupes de gènes orthologues. C'est important car cela suggère que c'est gènes/protéines appartenant au même groupe d'orthologues ont conservés la même fonction.

Sur l'arbre ComD, le groupe bleu est aussi un groupe de séquences orthologue. Les groupes verts sont issus d'une duplication car chaque sous arbre renferme au moins une séquence des espèces du groupe pyogenic et de *S. salivarius*. De plus, un des sous arbres verts renferme une paire de paralogues (Sube). Comme elles ne sont pas regroupées sous le même nœud, elles ne correspondent probablement pas à une duplication chez l'ancêtre de Sube. Nous pouvons remarquer, qu'une des deux copies a un orthologue dans Spar et que l'autre copie est directement en groupe externe de ces deux feuilles. Cette disposition suggère qu'une duplication a pu se produire chez l'ancêtre de Sube et Spar et qu'une des copies a été perdue chez Spar.

Le groupe rouge renferme trois séquences de Sgal. Deux pourraient correspondre à des duplications dans ce génome alors que l'autre est proche de Ssal. On remarquera aussi que les séquences de *S. thermophilus* (Sthe) et *S. salivarius*(Ssal) (groupe salivarius) ne sont pas regroupées et que les

séquences du groupe mitis sont retrouvées plus proche de Sequ que des autres séquences du groupe pyogenic. Ces positionnements inattendus pourraient provenir d'une mauvaise reconstruction de ces parties de l'arbre (hypothèse non soutenue par les valeurs de bootstrap qui sont relativement élevées dans cette région de l'arbre) ou de transferts horizontaux de gènes.

Les relations évolutives entre séquences inter groupes sont de type homologie ou paralogie ?

Tout dépend de l'hypothèse que l'on défend comme origine de ces groupes ! La présence de séquence appartenant au même génome dans les différents groupes suggère que les nœuds séparant ces groupes sont probablement associés à des évènements de duplication. Néanmoins, sous cette hypothèse, de nombreuses délétions indépendantes de gènes ont du se produire sur certaines branches, comme nous l'avons évoqué plus haut (hypothèse peu parcimonieuse !). Une autre hypothèse serait que ces groupes sont le résultat de transferts horizontaux qui auraient eu lieu très tôt chez les Streptocoques. Sous cette hypothèse deux gènes homologues d'un même génome ne sont pas issus d'un événement de duplication mais d'au moins une acquisition par transfert horizontal, on peut parler dans ce cas de pseudoparalogues (Koonin parle de pseudoparalogues http://www.cs.rice.edu/~nakhleh/COMP571/Presentations/Xiaoyun.pdf)

Que pensez-vous de la localisation des séquences ComE et ComD de *S. mutans* ?

Le système ComDE de *S. mutans* n'appartient pas au groupe bleu, il ne forme donc pas un système orthologue au système ComDE de *S. pneumoniae* et n'a donc probablement pas la même fonction. Ceci pourrait expliquer les différences de temps de latence observées lors de l'ajout du CSP avant le déclenchement de l'état de compétence chez *S. mutans*. Son système ComDE ne doit pas intervenir de la même manière dans la régulation de la compétence que le système ComDE de *S. pneumoniae*. Il est également possible qu'il ne soit pas impliqué dans cette régulation. C'est ce que nous savons aujourd'hui. La compétence étant régulée chez *S. mutans* par un autre système ComRS.

Régulation de la transformation par le système ComRS. On ne connait pas la régulation du niveau basal de l'opéron *comRS*, ce pourrait être un signal extracellulaire. Le produit du gène *comS*, le Pre-Coms, est exporté et maturé (ComS*) par un transporteur qui n'a pas encore été identifié. ComS* est importé dans la cellule par le transporteur d'oligopeptides Ami. Dans le cytoplasme, il se fixe à ComR et l'active. ComR activé se fixe sur les boites Com au niveau des promoteurs de *comS* et *comX*, conduisant à une amplification du signal (boucle auto catalytique) et à l'expression des gènes tardifs. Les protéines ClpC and MecA préviennent l'accumulation de ComX dans les conditions qui ne sont pas optimum pour le développement de la compétence.

Nous avons annoté les séquences en fonction de leurs domaines fonctionnels (HK et RR) et de leur appartenance à un système (un numéro). La reconstruction des systèmes repose sur la localisation des gènes sur le chromosome. Nous faisons l'hypothèse que les gènes proches appartiennent au même système.

Vous pouvez utiliser cette information pour analyser l'évolution des partenaires de chaque système. Qu'observez-vous ?

Si les deux arbres ne sont pas parfaitement congruents, nous observons globalement des topologies assez proches entre les HK et RR pour le groupe bleu et un des groupes verts. En termes évolutifs, cela indique que les deux partenaires du système, ComD et ComE, ont **coévolué**. Ainsi, dans le cas de transferts horizontaux, cela suggère que les deux gènes ont été transmis simultanément, ce qui est facilité par leur proximité chromosomique.